HighTech
FORUM

July 7, 2016

Ms. Marlene H. Dortch, Secretary
Federal Communications Commission
445 Twelfth Street, SW
Washington, DC 20054

Via Electronic Filing

Re: **WC Docket No. 16-106, Protecting the Privacy of Customers of Broadband and Other Telecommunications Services**

Dear Ms. Dortch,

I[1] offer these reply comments to aid the Commission in distinguishing factual technical claims offered in the privacy matter from false ones. Factual errors are at the root of the Commission's decision to apply Section 222 to the firms that it terms "Broadband Internet Access Providers" when it could forbear from applying these obviously telephone technology-centric regulations to Internet Service Providers.

The Commission may either press ahead with the regulations proposed in the NPRM or revise the proposed regulations so that ISPs and other parties in the Internet marketplace are governed by non-discriminatory, consistent regulations that promote higher quality advertising through increased competition. In either case, the Commission has an obligation to ensure that its rulemaking is based on a solid factual foundation.

FCC's Factual Errors

The NPRM is based on the erroneous belief that ISPs have greater access to consumer information than do other players in the Internet marketplace. This mistaken belief is copied into the Privacy NPRM from the Commission's "Open Internet Order" of February 26, 2015. The OIO declares:

> *Broadband providers serve as a necessary conduit for information passing between an Internet user and Internet sites or other Internet users, and are in a position to obtain vast amounts of personal and proprietary information about their customers [footnote].*[2]

[1] I am an independent network engineering consultant and policy analyst affiliated with High Tech Forum as founder and editor. These remarks are offered in my personal capacity and do not necessarily represent the opinions of any client or sponsor. I have previously offered comments in several FCC dockets, have offered testimony at the FCC En Banc Public Hearing on Broadband Network Management Practices in Cambridge on February 25, 2008 as an invited technical expert, and have testified before Congress on Internet privacy. My CV is available at http://www.bennett.com/resume.pdf.

[2] Federal Communications Commission, "Open Internet Order" (Federal Communications Commission, February 26, 2015), https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-24A1.pdf at ¶463.

The footnote refers to comments on deep packet inspection filed by human rights NGO *Access*: "See, e.g., Access Comments at 7 (stating that broadband providers have the technological capacity to exercise monitoring and control of their customers' use of the Internet using techniques such as deep packet inspection)."

The relevant portion of the Access comments is the following paragraph:

> To implement traffic management, ISPs often use tools with highly invasive capacities that can execute blocking, shaping, or filtering of data for unlawful political, social, and commercial purposes. These tools include deep packet inspection (DPI) technology. DPI allows ISPs - and anyone tapped into their networks - to identify and filter content while it traverses the internet, and make a copy of the traffic. DPI is the go-to mechanism governments across the world employ to invade user privacy and censor communications and content with staggering breadth and depth. In 2006, AT&T and the NSA were caught using DPI-capable technology in San Francisco to sort through all traffic flowing through a major switching station, in order to pick out specific messages based on targets like an e-mail address. Left unregulated, under paid priority schemes, ISPs will be incentivized to increase use of DPI to scour internet traffic in search of content to prioritize or degrade, down to the level of individual subscribers.[3]

The Access comments display a faulty technical understanding of the Internet, and incorrect grasp of the NSA's *Stellar Wind* program and a troubled relationship with fact and logic generally. The Access comments assert that ISPs routinely break the law; they confuse DPI with simple data replication (AKA "mirroring"); and they make an extraordinary connection between NSA surveillance and the bogeyman of the Open Internet Order, "paid prioritization."

Most importantly, these fear-mongering comments overlook the fact that Stellar Wind *aggregated* data mirrored by *multiple* ISPs and then *decrypted* that data in the agency's massive computer facilities. Without aggregation and decryption, no ISP has anything like the surveillance capability represented to the FCC by this misguided NGO. While it's understandable that the NGO would fail to grasp the facts, it's not acceptable for the FCC, an expert agency endowed with exceptional regulatory power, to accept this weak analysis as if it were factual.

Oddly, the privacy threat of greatest concern to Access is not commercial data gathering but NSA surveillance. By itself, this is an odd basis upon which to rely for justification for not forbearing from Section 222.

Code Breaking

NSA surveillance is accomplished in large part by a tool that ISPs lack, a comprehensive code-breaking capacity. In the case Access cites, AT&T, other ISPs, and transit providers mirrored packets passing through some of their optical switches to NSA, who performed

[3] Access, "Comment Re: Notice of Proposed Rulemaking on Protecting and Promoting the Open Internet" (Federal Communications Commission, July 18, 2014), https://ecfsapi.fcc.gov/file/7521700196.pdf.

the analysis, including decryption. Without the NSA's decryption capability, the potential for information gathering afforded to ISPs by virtue of their "position" in the Internet infrastructure is greatly diminished. And the FCC's privacy order does not appear to regulate the NSA.

The "position" to which the Access comments refer – by way of reference to an article Access cites from Wired Magazine – is not the position of the ordinary ISP.[4] Access refers to a surveillance operation known as Stellar Wind that collected data from telephone calls and email on peering links between AT&T and other telephone and Internet transit operators and Internet Exchange Points such as the Palo Alto Internet Exchange (PAIX) and Metropolitan Area Exchange, West (MAE-West).

Although the lawsuit filed by (my former co-worker) Tash Hepting against AT&T for its participation in Stellar Wind purported to represent the interests of AT&T's residential Internet users, Stellar Wind Internet data was not limited to AT&T's or Verizon's residential customers.[5] Participants in Stellar Wind were in the Internet transit business. This means that Stellar Wind participants had access to packets flowing between Internet users with no ISP business relationships with the firms who mirrored their packets to NSA.

These information packets were not analyzed or inspected by Stellar wind participants using any form of deep packet inspection. As the declaration of former AT&T technician Mark Klein in the lawsuit indicated, NSA got this data from mirrors attached to fiber optic links at the premises of the transit networks in question facing the Internet Exchanges.[6]

NSA's Unique Position

This is to say that the FCC relies on a representation by Access to impose telephone-era privacy regulations on ISPs – the claim that ISPs "are in a position to obtain vast amounts of personal and proprietary information about their customers" – when the NSA surveillance case that animates Access's concerns had nothing to do with ordinary Internet service or with the actual capabilities of ISPs.

In Stellar Wind NSA was "in a position to obtain vast amounts of personal and proprietary information" because it was able to draw upon data passing through not one but many ISPs and transit networks. In this respect alone – even if we overlook the encryption/decryption capacity of NSA – Stellar Wind was in a different position with respect to Internet traffic than is any individual ISP. Even if AT&T, Comcast, and Verizon were able to decode each information packet flowing through their networks,

[4] James Bamford, "The NSA Is Building the Country's Biggest Spy Center (Watch What You Say)," *WIRED*, March 15, 2012, https://www.wired.com/2012/03/ff_nsadatacenter/all/.

[5] Electronic Frontier Foundation, National Security Agency Telecommunications Records Litigation; Hepting v. AT&T (United States District Court Northern District of California 2009).

[6] Mark Klein, "Declaration of Mark Klein in Support of Plaintiffs' Motion for Preliminary Injunction" (United States District Court, Northern District of California, June 8, 2006), https://www.eff.org/files/filenode/att/mark_klein_unredacted_decl-including_exhibits.pdf.

these firms would only be able to see the information generated and requested by their own customers.

The reality of the Internet is that each edge service or application has the unfettered ability to see the data it exchanges with each of its customers. Google, Facebook, Amazon, and Netflix see customer-generated messages in plain text, after decryption. Similarly, these firms have unfettered access to the information they send to their customers before encrypting it.

There is a gap in the edge view of the Internet insofar as each edge service only sees information from its own customers. But this gap is reduced for advertising networks that are able to populate third party pages with ads. When an edge service operates both its own application and an advertising network – as many do – the gap becomes extremely small.

The ISP also has a limited view of the Internet for three reasons:

1. Each ISP can only see information generated or received by its own customers;
2. Most of this customer data handled by the ISP is encrypted; and
3. The data the ISP can see is devoid of context.

The first limitation is shared by ISPs, edge services, and advertising networks insofar as each can only view data exchanges involving its own customers. But this factor argues for regulating ISPs less heavily than the large edge and ad companies because the number of users each ISP has is much smaller than the corresponding number in the edge and ad space. The largest wireline ISP, Comcast, has 23 million customers.[7] Netflix has 81 million customers worldwide;[8] Amazon had 244 million users in 2014;[9] Facebook has 1.59 billion customers;[10] Google has seven different services with over a billion users *each.*[11] There is no dearth of advertising-relevant data for edge services to capture and use. For ISPs to catch up in terms of user counts, each would need to grow by one to two orders of magnitude, signing up more Internet users than the planet contains.

As noted, the increased use of encryption reduces the value of the customer data passing through ISP facilities. Unless ISPs are willing to invest in an NSA-caliber code-breaking facility, the only elements of sensitive transactions visible to ISPs are destination IP address, data volume, application type (e.g., web page vs. video stream vs. phone call), and transaction time of day. DNS lookups duplicate IP addresses and can be exported to

[7] Jon Brodkin, “Comcast Shrugs off Years of Cord-Cutting Losses, Adds 89K TV Customers,” *Ars Technica*, February 3, 2016, http://arstechnica.com/business/2016/02/comcast-shrugs-off-years-of-cord-cutting-losses-adds-89k-tv-customers/.

[8] “Netflix : Overview,” accessed July 7, 2016, https://ir.netflix.com/.

[9] “How Many Customers Does Amazon Have? -- The Motley Fool,” accessed July 7, 2016, http://www.fool.com/investing/general/2014/05/24/how-many-customers-does-amazon-have.aspx.

[10] “Here’s How Many People Are on Facebook, Instagram, Twitter and Other Big Social Networks,” accessed July 7, 2016, http://adweek.it/1qqjteI.

[11] “Google Has 7 Products With 1 Billion Users | Popular Science,” accessed July 7, 2016, http://www.popsci.com/google-has-7-products-with-1-billion-users.

third party DNS providers in any case (and would be, if such activities were truly valuable).

The context factor is extremely significant and often overlooked. Raw packet streams contain more noise than signal, while application transactions take place in a coherent context. When we perform an Internet search, the search engine knows we're doing a search and which search terms we use without expending any processing resources to speak of.

Examining a stream of packets to determine the same information is much more processing-intensive when it can be done at all; Google, Bing, and Yahoo encrypt Internet searches. But even if they didn't, extracting searches from IP packets does indeed require DPI, a considerable expenditure of processing power.

Creating context for social network interactions from raw packet streams also requires a great deal of processing that isn't required by the social network itself. The business of the social network is all about keeping track of the people and subjects that attract our interest. For an ISP to develop the dossiers social networks maintain on users would require at least an equal expenditure of processing power as that expended by the social network in addition to the processing power necessary for the ISP to conduct its business as an ISP. And the ISP would need to apply this processing power in a different way for each edge service whose interactions it wanted to track. This is probably unrealistic.

Consequently, ISPs are not the NSA and they don't have the ability to comprehensively survey every – or even many – of the transactions that take place over the Internet.

Conclusion

Like the game of *Telephone*, the facts of Stellar Wind are distorted by the Hepting/EFF lawsuit, further twisted by the *Wired* article, misrepresented by Access, misconstrued by the FCC's Open Internet Order and confused again by the FCC's Privacy NPRM. ISPs do not have the surveillance advantage over edge services and advertising networks the NPRM attributes to them.

Consequently, the privacy NPRM lacks a coherent factual foundation for the claim that ISPs must be regulated differently than edge services because of their unique vantage point in the Internet.

In reality, edge services, browsers, operating systems, advertising networks, and transit networks all have better and more comprehensive knowledge of user interactions with edge services than ordinary ISPs do. As this is the case, the FCC's decision to impose Section 222 with a new set of rules deeply at odds with the FTC Privacy Framework is irrational.

The more prudent course is to forbear from imposing the Section 222 opt-in provision on Internet service providers and to generally harmonize ISP privacy regulations with the FTC framework. Opt-in is appropriate for sensitive information but not for generic interactions.

Supporting Material

The following includes recent blog posts pertinent to the Privacy NPRM as well as written testimony I gave to the House Energy and Commerce Communications Technology and the Internet Subcommittee on Internet privacy in April, 2009.

The blog posts provide technical analysis of the current state of privacy in the Internet and of the debate about Internet privacy policy.

This testimony precedes my employment in the public policy field. The testimony discusses the origin and use of deep packet inspection tools and offers a comprehensive picture of threats to Internet privacy. The testimony is available online at HighTechForum.org http://hightechforum.org/wp-content/uploads/2016/07/Privacy-Testimony-2009.pdf

The blog posts are first and the Congressional testimony follows.

Appendix A: Bringing Privacy Into the Open

This High Tech Forum blog post from January 26, 2016 addresses a letter sent to the FCC by a group of privacy advocates. It is available at http://hightechforum.org/bringing-privacy-into-the-open/

Bringing Privacy into the Open

We don't write about Internet privacy a great deal because it's much more a matter of pure policy than one of technology. Privacy policy has more to do with consent, retention, protection, and dissemination than with how computers and networks actually work, so most of it is outside the scope of a technology blog such as this one. But there are some important technical issues involved in privacy that come to the surface when legislators and regulators consider how to tailor privacy policy to the various players in the Internet space. As we see in so many policy battles, groups of firms aligned by business models fight to protect their common interests from regulation while seeking to impose harsh restrictions on firms with different business models.

The first time I testified before Congress, back in 2009, the topic was privacy. Advertisers and advertising networks are the obvious candidates for privacy regulation because their businesses get a boost from detailed knowledge of user preferences. At the most basic level, advertisers don't want to waste money showing Lexus ads to poor people or showing ads for fancy women's shoes to teenaged males. In principle, the better the targeting the more advertisers are willing to pay. A lead sheet is more valuable than a billboard.

At the Congressional hearing, advertisers pushed back on the Internet subcommittee's desire to regulate their industry by claiming ISPs had access to more information than they did. While this was true in a superficial sense – all the packets going to Amazon, Google, and eBay pass through an ISP at some point – it's also disingenuous because ISPs have a business model that depends on subscription fees rather than ad sales. So even if your ISP knows the details of every transaction you make on the Internet, the information doesn't translate into revenue.

Or at least it didn't until recently. ISPs have discovered that they can reduce the prices they need to charge users if they can sell some information to ad networks. This allows the ISPs to connect the data they're in a position to harvest with a means of using it to pitch products to consumers, something the ISPs can't do on their own since they can't alter the information that goes to your web browser. Google taught this lesson to the ISPs by offering extremely low cost, very high speed Internet service so that they can harvest more information to use in their advertising auctions. AT&T notably followed by matching Google's price – $70 for a gigabit pipe – on the condition that they can harvest personal information and sell it to advertisers. The same service is available with privacy for $100 a month. Most people choose the $70 plan, of course.

Differential privacy regulation got a boost when the FCC reclassified Internet access under Title II of the Communications Act because that move split privacy between the FCC and the Federal Trade Commission according to Internet business models. The FCC controls privacy for Title II businesses and the FTC has the ball for advertisers and web sites. Consequently, the advertisers are now pressing the FCC to impose severe restrictions on ISP use of personal information while arguing before the FTC that industry self-regulation is the way to go.

This double standard approach came to a head last week when a collection of self-styled privacy advocates presented a letter to the FCC arguing that ISPs are uniquely positioned to spy on consumers and hence must be harshly controlled:

> *Providers of broadband Internet access service, including fixed and mobile telephone, cable, and satellite television providers, have a unique role in the online ecosystem. Their position as Internet gatekeepers gives them a comprehensive view of consumer behavior and until now privacy protections for consumers using those services have been unclear. Nor is there any way for consumers to avoid data collection by the entities that provide Internet access service.*

While ISPs are first and last to carry bits between consumers and Internet services, the claims about "gatekeeping" and the helplessness of consumers to guard personal information from ISPs are factually challenged. There are many gates and many gatekeepers on the Internet, of course. Many transactions begin with a Google search, many gaming sessions take place over Facebook, and many purchases are mediated by Amazon, eBay, or PayPal. Payment services have the most luscious information of all: what we purchase, where we purchased it, and how much we paid for it. That's not exactly small potatoes, and it makes more sense to use PayPal with its top-notch security than some tiny web site of uncertain reputation to handle your credit card information.

Peter Swire, the privacy czar in the Clinton Administration, pushed back on the privacy hawk letter with a technical analysis of the two main claims, which he finds technically dubious for the same reasons I do. In the first place, many web services now use HTTPS, which encrypts the information exchanged between user and web site. If I search for "cats" on Google, the URL looks like something like this:

> https://www.google.com/search?q=cats&rlz=1C5CHFA_enUS563US566&oq=cats&aqs=chrome.0.69i59j0l5.2170j0j9&sourceid=chrome&es_sm=91&ie=UTF-8

This causes the browser to do a DNS query for www.google.com and to send the rest of the query in encrypted format to the IP address returned by DNS. Hence, the only information the ISP can see is the destination IP Address and its equivalent domain name. So the degree of access the ISP has is determined largely by the service that's being used (Google in this case.) A lot can be made of the fact that the DNS query is done on plain text, but that's about to change.

IETF has a DNS Privacy working group developing a means of cloaking DNS queries going to external resolvers operated by Google and a few others; it's called "DPRIVE." DPRIVE started last summer with an informational document, RFC 7626, setting out goals and principles. It addresses some of the myths about DNS:

> *2.1. The Alleged Public Nature of DNS Data*
>
> *It has long been claimed that "the data in the DNS is public". While this sentence makes sense for an Internet-wide lookup system, there are multiple facets to the data and metadata involved that deserve a more detailed look. First, access control lists and private namespaces notwithstanding, the DNS operates under the assumption that public-facing authoritative name servers will respond to "usual" DNS queries for any zone they are authoritative for without further authentication or authorization of the client (resolver). Due to the lack of search capabilities, only a given QNAME will reveal the resource records associated with that name (or that name's non-existence). In other words: one needs to know what to ask for, in order to receive a response. The zone transfer QTYPE [RFC5936] is often blocked or restricted to authenticated/authorized access to enforce this difference (and maybe for other reasons).*
>
> *Another differentiation to be considered is between the DNS data itself and a particular transaction (i.e., a DNS name lookup). DNS data and the results of a DNS query are public, within the boundaries described above, and may not have any confidentiality requirements. However, the same is not true of a single transaction or a sequence of transactions; that transaction is not / should not be public. A typical example from outside the DNS world is: the web site of Alcoholics Anonymous is public; the fact that you visit it should not be.*

This puts the focus on cloaking the lookup, and the working group is developing the means to do that. When DNS lookups are cloaked, the ISP loses its special power to know which domains you're interested in visiting.

Next, you can also cloak web queries – the http stuff – and off-domain references by using VPNs, as many people do when working from home. VPNs have a performance penalty, so not that many people use them because they can make pokey web sites even pokier. They also have a cost, which is really what this battle is all about.

Companies that are heavily invested in advertising already provide DNS service to keep destinations hidden from ISPs, and the same firms that do that can easily provide VPNs for attractive prices – like free – for the same purpose. The fact that they don't currently do this does not change the fact that the second claim in the privacy letter is false. It is simply not the case that "consumers [cannot] avoid data collection by the entities that provide Internet access service." VPNs are such a means. They aren't popular, but they exist. In many countries, VPNs are essential to using Netflix, which is a bit embarrassing as they're a means of bypassing content licenses.

If DNS as it currently works is a privacy vulnerability, this says something interesting about the nature of the service offered by ISPs, as a matter of fact. And the fact that DNS can be provided by third parties also says something interesting about differential privacy regulations.

As the FCC sees it, collecting information from DNS queries answered by an ISP is regulated under Title II because DNS is inseparable from the "offer" of Internet service. But the FCC's logic also says that collecting information from DNS queries submitted to Google is exempt from Title II and FCC jurisdiction because Google is not an ISP (except in a few towns with an unknown number of users.)

Does that make sense? If the FCC and the FTC adopt uniform regulations for DNS privacy the double standard goes away, but that will take us into the territory where the FCC's Title II regulations extend beyond the scope of ISPs and into the fabric of the Internet itself.

Perhaps this slippery slope is inevitable, but it raises a host of questions.

Appendix B: CDT's Diagram Muddies the Waters

This High Tech Forum blog post from February 9, 2016 addresses a memo from CDT on the issues facing the FCC as it begins a privacy rulemaking. The memo conveyed errors of fact and analysis that I sought to correct. It is available at http://hightechforum.org/cdts-diagram-muddies-the-waters/

The week before last, I wrote about the factual errors in a letter to the FCC signed by a collection of advocacy groups with an interest in consumer privacy. Briefly, the letter grossly overstated the amount of personal information available to ISPs from Internet communications and understated the ease with which consumers can shield communications from the (prying?) eyes of ISPs. The reality is that many Internet "edge services" such as Amazon and Google can and do encrypt communications with their users and customers, and consumers are free to use Virtual Private Networks to cloak communications that would otherwise be unencrypted.

One of the signatories is the Center for Democracy and Technology (CDT,) a group that generally has a better grasp on Internet technology than the other signatories, Free Press and Public Knowledge in particular. CDT employs actual technologists, which can't be said for the others. But CDT is mainly a group of lawyers, so we have to take their technical analysis with a bit of context.

Two of the CDT lawyers, Alex Bradshaw and Stan Adams, have written a blog post on the issues facing the FCC as it seeks to apply its new Section 222 authority to the Internet. Most of the post reads like this:

> *Sections 222(c) and (d) control how, and under what conditions, carriers may use, share, or disclose "Customer Proprietary Network Information" (CPNI), the definition of which includes information related to the "quantity, technical configuration, type, destination, location, and amount of use of a telecommunications system…and information contained in bills."*

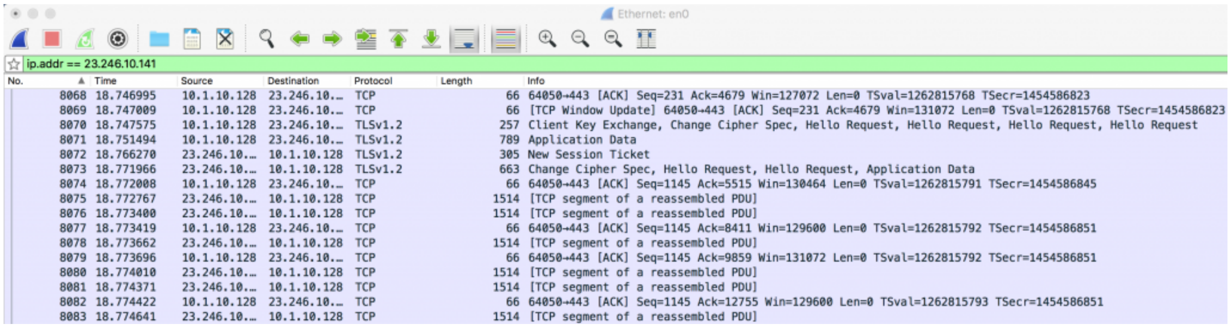
This probably means that ISPs must obtain customer consent before sharing information about Web surfing habits with ad networks, or it would mean that if the ISPs were on a level playing field with Amazon and Google. The lawyers will sort that out.
The blog post is accompanied by a memo that attempts to decompose Internet services into component parts that adds nothing to the analysis in the CDT blog, and actually detracts from it by containing the same kinds of nagging errors Peter Swire found in the FCC letter. The memo argues that application level encryption leaves protocol headers unencrypted, and seems to argue that unencrypted protocol headers (for TCP and IP) provide ISPs with important information about what the customer is doing. As the memo puts it:

> *Long-term monitoring of packet headers traveling to and from IP and MAC addresses can reveal patterns and associations that paint a picture of what kinds of information customers are sending and receiving over the Internet, and when, where, and how they do so. For instance, by looking at packet size, packet streams, and IP addresses, a network operator could infer that you are streaming a movie from a particular content provider. A network operator could begin to develop a comprehensive profile of your broadband usage patterns, or even your personal habits, like when you sleep, work, watch movies and send email.*

While this is at least partially true, it has very little significance. The encrypted packets that pass between Google search and a Google user simply expose the fact that an interaction takes place, not the identity of the user or the terms the user is looking for. It could be anyone in the house, even a visitor, and the search can be about anything. Only Google knows for sure who submitted the search request or what it was about.

Streaming a movie from Netflix provides no more information than a Google search, since it's encrypted in TLS v1.2. The ISP can find out that someone in the house initiates a Netflix streaming session, but does not know which person and which movie. Again, the specific identity of the user and the content is known only to Netflix.

I verified this by examining a Netflix session with Wireshark, a network analyzer. This is what the capture looks like. The lines labeled "TLSv1.2" mean the connection is encrypted, which is confirmed by the fact that the TCP packets contain gibberish.



```
ip.addr == 23.246.10.141
No.      Time       Source       Destination  Protocol  Length  Info
   8068 18.746995  10.1.10.128  23.246.10.…  TCP         66 64050→443 [ACK] Seq=231 Ack=4679 Win=127072 Len=0 TSval=1262815768 TSecr=1454586823
   8069 18.747009  10.1.10.128  23.246.10.…  TCP         66 [TCP Window Update] 64050→443 [ACK] Seq=231 Ack=4679 Win=131072 Len=0 TSval=1262815768 TSecr=1454586823
   8070 18.747575  10.1.10.128  23.246.10.…  TLSv1.2    257 Client Key Exchange, Change Cipher Spec, Hello Request, Hello Request, Hello Request, Hello Request
   8071 18.751494  10.1.10.128  23.246.10.…  TLSv1.2    789 Application Data
   8072 18.766270  23.246.10.…  10.1.10.128  TLSv1.2    305 New Session Ticket
   8073 18.771966  23.246.10.…  10.1.10.128  TLSv1.2    663 Change Cipher Spec, Hello Request, Hello Request, Application Data
   8074 18.772008  10.1.10.128  23.246.10.…  TCP         66 64050→443 [ACK] Seq=1145 Ack=5515 Win=130464 Len=0 TSval=1262815791 TSecr=1454586845
   8075 18.772767  23.246.10.…  10.1.10.128  TCP       1514 [TCP segment of a reassembled PDU]
   8076 18.773400  23.246.10.…  10.1.10.128  TCP       1514 [TCP segment of a reassembled PDU]
   8077 18.773419  10.1.10.128  23.246.10.…  TCP         66 64050→443 [ACK] Seq=1145 Ack=8411 Win=129600 Len=0 TSval=1262815792 TSecr=1454586851
   8078 18.773662  23.246.10.…  10.1.10.128  TCP       1514 [TCP segment of a reassembled PDU]
   8079 18.773696  10.1.10.128  23.246.10.…  TCP         66 64050→443 [ACK] Seq=1145 Ack=9859 Win=131072 Len=0 TSval=1262815792 TSecr=1454586851
   8080 18.774010  23.246.10.…  10.1.10.128  TCP       1514 [TCP segment of a reassembled PDU]
   8081 18.774371  23.246.10.…  10.1.10.128  TCP       1514 [TCP segment of a reassembled PDU]
   8082 18.774422  10.1.10.128  23.246.10.…  TCP         66 64050→443 [ACK] Seq=1145 Ack=12755 Win=129600 Len=0 TSval=1262815793 TSecr=1454586851
   8083 18.774641  23.246.10.…  10.1.10.128  TCP       1514 [TCP segment of a reassembled PDU]
```

So the memo and its accompanying diagram are a lot of smoke and mirrors. ISPs have very limited insight into what individual subscribers are doing across the Internet; packets are encrypted, protocol headers provide very little information due to the fact that everyone on a household shares a common IP address (the one that belongs to the router) and because individual users do not share MAC addresses with the ISP. Hence, CDT's initial claim that ISPs can monitor IP and MAC (Ethernet and Wi-Fi interface addresses) is false.

We access the Internet from our home routers, which contain a function known as a Network Address Translator that replaces the IP and MAC addresses of the devices within the home with a common, global IP address and the MAC address the router port that connects to the DSL or cable modem. To get information about who is doing what you need to be on the other end of the Internet, where Google and Netflix sit.

CDT's description of packets, protocols, layers, and encryption is nice, but they should look at some actual data flows before jumping to conclusions. The post does raise an interesting question, however: Assuming that ISPs and edge services should be on a level playing field with respect to privacy, if it's troubling that ISPs might be able to decode such patterns as "your broadband usage patterns, or even your personal habits, like when you sleep, work, watch movies and send email," is it equally troubling that Google, Amazon, and Netflix might be able to profile user information that happens to be at least that personal?

Is it troubling that Amazon knows I use Tom's toothpaste, Google knows I'm taking an overseas trip this week, and Netflix knows I like to watch cowboy cop shows like "Longmire?" It doesn't trouble me a great deal that anyone knows these things because I've just disclosed them, maybe even truthfully.

Finally, CDT's decomposition of Internet data formats is cloudy because it omits the most important element of Internet use, the stream. Packets, protocols, layers, and headers are much less important that the thing they enable, which is streams of data between users and users or users and services. Information streams are the most important element of Internet interaction, and any analysis of the Internet that fails to mention them isn't very useful. Streams will be the subject of a post to come shortly.

Published at High Tech Forum.

Appendix C: FCC Confused about Privacy

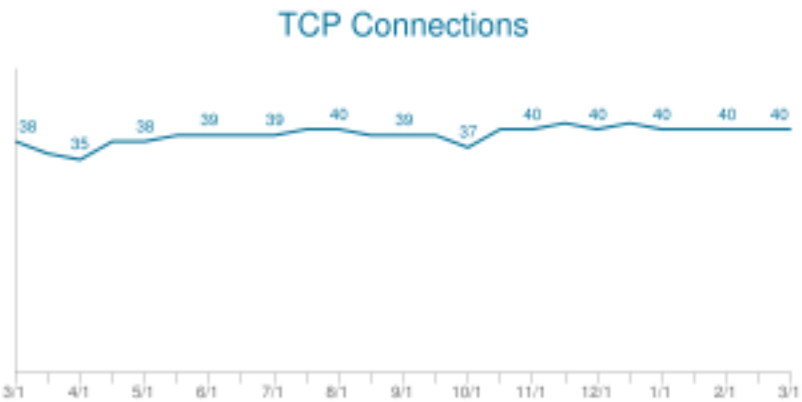
This High Tech Forum blog post from March 10, 2016 addresses the vantage point error drawn into the Privacy NPRM from the OIO. It is available at http://hightechforum.org/fcc-confused-about-privacy/

FCC Confused About Privacy
As we've explained in previous posts, the FCC is often very confused about the way the Internet works. The agency has claimed that the Domain Name Service routes information across the Internet, for example, when it really does little more than translate domain names into IP addresses. The routing of packets is handled by the Internet Protocol (IP) itself, with help from a protocol known as Border Gateway Protocol (BGP) that allows IP routers to build maps of the Internet.

FCC Chairman Wheeler's ISP privacy proposal makes similar errors, and adds a new logical error that wasn't quite so evident in his "open Internet" rulemaking.
The factual error is the claim that consumers are powerless to hide their Internet activities from the presumptively untrustworthy ISPs. The proposal falsely claims that consumers necessarily share web surfing details with ISPs:

Even when data is encrypted, broadband providers can still see the websites that a customer visits, how often they visit them, and the amount of time they spend on each website.

This claim is partially true for consumers who simply visit sites like Google.com that are encrypted by default with TLS, but it's not at all true for consumers who use VPNs. It's only partially true because the information that TLS exposes to ISPs is limited to IP addresses and flows to and from those addresses. This information is a lot less useful than Wheeler imagines because web pages are composed of page elements that have their own IP addresses and data flows. This is easy to confirm by looking at the ads that accompany typical web pages.



Number of TCP Connections per Web Page. Source: HTTP Archive.

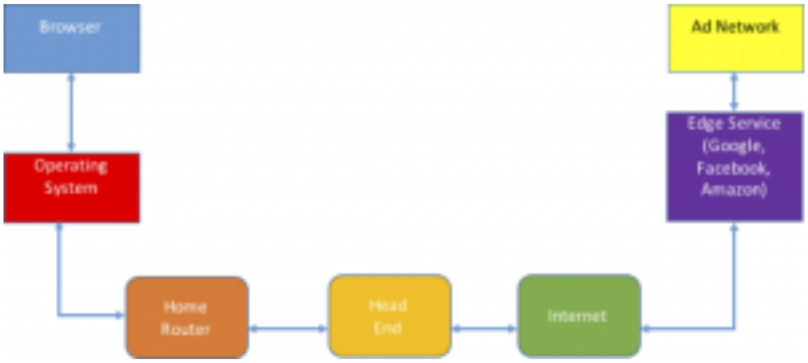
These ads are served up by ad servers associated with ad networks, and there can be cases in which more actual packets come to the user from ads than from the substance of the web pages he or she visits. This is especially true for the annoying video ads that accompany so many web pages these days, especially pages devoted to sports and news. The typical web page contains 54 image requests today, which account for 1.5 MB of data.

The HTML of the web page itself averages a mere 67 KB by comparison. All told, the typical web page entails 40 different IP addresses today, and all the ISP can do with the all that information is guess what the important parts are and what parts are merely ads. It would require an enormous amount of computation to figure out what the user is doing at these IP addresses and which, if any, are actually interesting to the user.

As a practical matter, converting the raw information that ISPs can harvest from web requests made by users who aren't using VPNs is a very difficult task. The information is both highly random and very fragmented. So Wheeler's proposal assumes a status quo that doesn't really exist.

This is par for the course in the public commentary around the privacy issue from outside the FCC as well. Professor Nick Feamster has written letters to the FCC as well as blog posts that claim ISPs are in a position to see "much more user traffic from many more devices than other parties in the Internet ecosystem…"

This is factually incorrect for two reasons and misleading in one other. Even though ISPs may be "in a position" to see a lot of traffic doesn't mean they actually do anything with the traffic they can see or even that seeing the traffic is the same as understanding it. In fact, it's a lot harder to extract any user preference information from a raw data stream – even when it's unencrypted – than it is to extract useful data on the other side of the web.



Internet Vantage Points

Let’s look at a little diagram of the Internet’s vantage points where web surfing is concerned. Users enter URLs in their web browsers, which pass web page requests to the TCP element in their operating system (Windows or MacOS.) The operating system passes it on to the home router, which sends it up to the headend of the cable modem network (or its equivalent in non-cable networks.) The ISP network connects to the web server, either directly or through a transit network. If the web server is commercial in any way, it relies on a series of advertising networks to record the user’s visit and offer up suggestions on the ads the user is going to see.

The ownership of many of these vantage points varies. Consumers use multiple web browsers, operating systems, and home routers. Web servers use a variety of transit networks and ad servers. The fact that home routers can come from either ISPs or the consumer’s retail choice is the largest omission in the Feamster analysis. He assumes the home router is the primary vantage point for the prying ISP, which completely falls apart when users buy their own routers and simply use the ISP equipment to carry packets without awareness of which user in the house is looking at the given web page.

Both Feamster and the FCC make the gigantic leap from the rather limited fact that ISPs know the IP addresses of all the non-VPN packets they carry to the conclusion that they can actually do something with this information. There is a strangely inconsistent application of what ISPs and consumers might be able to do with what they actually do. When discussing consumers, the tendency of privacy advocates and the FCC is to disregard the protections of personal privacy afforded to those who exercise opt-out rights and use VPNs because most consumers don’t care very much about privacy. If the tools are available to consumers but consumers don’t use them, it doesn’t matter what they tell pollsters because their behavior says they don’t care about Internet privacy.

But when discussing ISPs, the balance of “can” and “do” swings in the opposite direction. Even though ISPs are capable of adding surveillance code to home routers, there’s no evidence that they actually do. When I made home routers for ISPs I didn’t get a single request to track web site visits. If I had, it would have been an enormous project that would have doubled or tripled the price my company charged our ISP customers. And we had control of the code in the home router, not the ISPs.

So what difference does it make that the ISPs *could* track user Internet behavior if they *don't*?

As to the claim that ISPs are in a position to see much more user traffic than other parties in the Internet ecosystem, this certainly isn't true in relation to advertising networks. For encrypted web sites, only the web servers and ad networks know which pages the user is visiting. This information is much more valuable than simply having a collection of dozens of IP addresses for each web page the user visits and no way to make any sense of them.

So the Wheeler and Feamster analyses seem to express a concern that ISPs may someday develop the capability to parse user activities that can rival and possibly even surpass the capabilities that ad networks already have today. Therefore, the FCC's privacy inquiry compares the *possibility* of the ISPs becoming serious rivals to the ad networks to the *reality* that ad networks have more information than any other party in the Internet space regarding user web activity.

I'm not sure that would be such a bad thing. But even if it is, shouldn't we be talking about ISP data collection practices versus those of ad networks if we want to have a coherent policy dialog? The false claims, misdirection, and cherry-picking about who knows what about whom is preventing this discussion from taking place, and that's unfortunate.

Published at High Tech Forum

Appendix D: Internet Architecture vs. Section 222

This High Tech Forum blog post from June 10, 2016 addresses the fundamentally different architectures of the Internet and the telephone network and how those differences way on the expectation of and responsibility for privacy. Available at http://hightechforum.org/internet-architecture-vs-section-222/

Internet Architecture vs. Section 222

If you follow debates about Internet policy in detail, you will often find advocates arguing opposite sides of particular questions in different contexts. Responsible advocates avoid this practice because, obviously, it undermines their credibility because astute observers notice it. But we don't simply conduct these debates among well-informed and thoughtful participants any more. The pop culture audience is increasingly important in Internet policy because so many people have a stake in the ongoing health and welfare of the Internet, and also because the general audience can, when aroused by advocates, flood the FCC and Congress with canned letters of protest. They also generate traffic to tech policy blogs, especially those that deal with the debates in an emotional, manipulative way.

Net Neutrality vs. Net Neutrality

Some remarks from Public Knowledge about the subject of encryption in the context of Internet privacy are sharply opposite arguments that organization made in the net neutrality debate. The rationale for passing net neutrality regulations comes down to a desire to protect the underlying design of the Internet – its architecture – from meddling by ISPs and other potentially bad actors. This architecture has been claimed by net neutrality advocates to come down to the so-called "end-to-end arguments principle" that holds that new features and functions needed by Internet applications should be provided by the applications themselves and not by actions taken by the Internet Service Providers, the network equipment vendors, or the long-haul transit networks themselves.

Net neutrality advocate Barbara van Schewick wrote a 600 page book on this one idea, Internet Architecture and Innovation. Amazon summarizes it in the following way:

> *The Internet's original architecture was based on four design principles: modularity, layering, and two versions of the celebrated but often misunderstood end-to-end arguments. But today, the Internet's architecture is changing in ways that deviate from the Internet's original design principles, removing the features that have fostered innovation and threatening the Internet's ability to spur economic growth, to improve democratic discourse, and to provide a decentralized environment for social and cultural interaction in which anyone can participate. If no one intervenes, network providers' interests will drive networks further away from the original design principles. If the Internet's value for society is to be preserved, van Schewick argues, policymakers will have to intervene and protect the features that were at the core of the Internet's success.*

End-to-End Design Principles

The end-to-end arguments principle was originally articulated in a paper by two MIT post-docs and their supervisor, Jerry Saltzer, titled "End-to-End Arguments in System Design" published in 1981. Oddly, the paper does not contain the word "Internet" but advocates argue that it describes the architecture that motivated its design nonetheless. Of particular interest is the one paragraph on security by encryption:

> *The end-to-end argument relating to encryption was first publicly discussed by Branstad in a 1973 paper[2]; presumably the military security community held classified discussions before that time. Diffie and Hellman[4] and Kent[8] develop the arguments in more depth, and Needham and Schroeder[11] devised improved protocols for the purpose.*

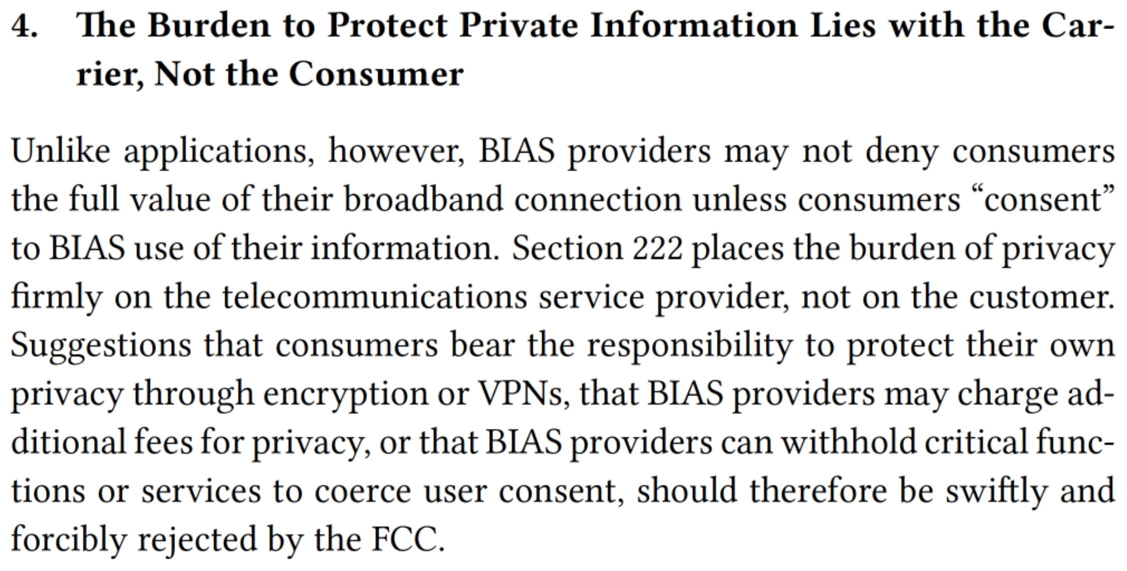
This is to say that even before there was an Internet it was recognized in computer science that networks alone cannot provide end-user security unless applications take on the responsibility of encrypting information. Regardless of which bad actor you're worried about, the best way – and indeed the only way – to protect confidential communication is to encrypt your messages and nobody can do this for you.

With Great Power Comes at Least Some Responsibility

So end-to-end architecture means that users and applications have power they don't have on centralized networks like the old-school public switched telephone network. But as the mighty American philosopher Spiderman told us, "with great power there must also come — great responsibility." So it's up to users and applications to step up and take affirmative steps to protect their communications from abuse. Users of the Internet are also responsible for protecting their devices from viruses by using malware programs, avoiding dodgy websites, and practicing good password hygiene. Even if you practice safe surfing you may be hacked anyway because no one is really immune, but it's your obligation to try in any event.

Encryption hasn't always been practical because it requires CPU power, but today's computers have enough performance that we can and therefore we should. And this includes avoiding websites that don't use TLS encryption (https) just as you would avoid financial institutions whose websites don't require annoying two-factor authentication rituals before you can transfer money out of them.

"Oh no, we didn't mean THAT End-to-End idea"
But Public Knowledge disagrees with the idea that Internet users are responsible for protecting their communications from snooping after spending much of the last decade swearing allegiance to the Internet's end-to-end architecture. This is what they wrote in their white paper, *Protecting Privacy, Promoting Competition: A Framework for Updating the Federal Communication Commission Privacy Rules for the Digital World:*



4. The Burden to Protect Private Information Lies with the Carrier, Not the Consumer

Unlike applications, however, BIAS providers may not deny consumers the full value of their broadband connection unless consumers "consent" to BIAS use of their information. Section 222 places the burden of privacy firmly on the telecommunications service provider, not on the customer. Suggestions that consumers bear the responsibility to protect their own privacy through encryption or VPNs, that BIAS providers may charge additional fees for privacy, or that BIAS providers can withhold critical functions or services to coerce user consent, should therefore be swiftly and forcibly rejected by the FCC.

Public Knowledge correctly observes that telephone network regulations placed the "burden of privacy" on the network rather than the user but they fail to comprehend why this was the case. It's not very difficult to understand even if you're not familiar with end-to-end architecture: For most of the PSTN's life, there was no practical way for the end user to protect his or her own privacy because all the equipment between calling party and called party was owned and operated by the telephone service provider. Users could speak in code, but few did unless they were up to something because both parties had to be up on the same convention for that to work.

It's Up to You, Gentle Surfer

So if you're interested in enjoying full privacy on the Internet it's not only your responsibility, it's a power the Internet gives you because of the way it's designed. And sure enough, encrypting messages is not only something that protects you from bad actors like the Russian mafia: it also protects you from good actors such as your ISP, your search provider, or your email service when they make mistakes and expose your communication to hackers without meaning to.

In reality, the notion that ISPs can protect your privacy without your making an effort to meet them halfway is fiction. All the ISPs can do with respect to "privacy" is refrain from competing with applications such as search, email, and web surfing for the sale of your information to advertisers. That refusal is neither privacy-enhancing to users nor is it pro-competition.

Regulation that requires ISPs to "protect your privacy" by refraining from paying attention to what you're doing is also inconsistent with end-to-end architecture. The argument that ISPs must refrain from trafficking in the raw material of advertising sales in order to comply with the literal text of Section 222 also tells us that there's a major disconnect between PSTN regulation and the goals and purposes of the Internet. But you already knew that.

The so-called privacy debate is really a debate about the Internet advertising market. It's best to have this discussion in a way that doesn't offend the principles that the net neutrality battle was supposed to protect, but that is not happening.

Published at High Tech Forum.

Appendix E: Congressional Testimony, April 2009

This is the Congressional testimony I offered on Internet privacy in April 2009 while working as a software engineer in Silicon Valley. I include it here because the copy on the House web server suffers from link rot. A copy is available at http://hightechforum.org/wp-content/uploads/2016/07/Privacy-Testimony-2009.pdf

April 21, 2009

The Honorable Chairman Rick Boucher
The Honorable Ranking Member Cliff Stearns
The Honorable Chairman Henry A. Waxman
The Honorable Ranking Member Joe Barton
The Honorable Members of the Subcommittee on Communications, Technology, and the Internet

Subject: Testimony before the House Committee on Energy and Commerce Subcommittee on Communications, Technology, and the Internet hearing on April 23, 2009

Dear Chairman Boucher, Ranking Member Stearns, and members of the Subcommittee,

Thank you for offering me the opportunity to address the subcommittee on the subject of technologies that monitor consumer use of communications networks. The topic is pertinent to the evolution of the networks, to the development of consumer awareness, as well as to potential new regulations if they're needed. I'd like to offer a few recommendations.

Network Monitors

I'm a network engineer, inventor, and writer. I've designed data transfer and Quality of Service protocols for some of our most widely used communications networks, switched Ethernet and Wi-Fi, as well as for some that haven't got off the ground yet, such as Ultrawideband. As a consequence, I've had occasion to use a variety of network monitoring and analysis equipment to observe traffic on networks.

Network monitors – often called "Sniffers" after a popular product produced by Network General in the 1980s – enable engineers to see every part of the packets that traverse the network segments to which the monitors are attached, including the various payloads present at the Ethernet, IP, TCP, HTTP, and Content layers, for example. These are vital tools that permit programmers and electrical engineers to accelerate systems and isolate and correct bugs that would otherwise limit network function. These systems pre-date the Internet by many years, and it's safe to say that we would have no working packet networks without them.

These devices have political uses as well – the controversy over Comcast's first-generation traffic shaping system was set off by a network technician who used an open source network monitor to discover suspicious TCP packets mingled in amongst the peer-to-peer file sharing packets he expected to see. Since the 1980s, these devices have had the ability to apply filters to network traffic based on sophisticated pattern matching, to produce logs of selected packets, and to perform a variety of statistical analyses to network traffic streams. They are frequently used by network administrators to troubleshoot problems in both local and wide-area networks, and are generally considered to be invaluable aids in maintaining the semblance of stability that users expect of their networks.

While these monitors have been used on occasion to steal passwords and other user information, these instances are rare and limited in scope simply because an Ethernet monitor can only be used to capture traffic on the particular part of the network to which it is attached. If I monitor the traffic on my home network, as I frequently do, I can't see any of the traffic generated by my neighbors, even though we share a common coaxial cable to a shared CMTS; this is because my cable modem only passes traffic intended for my Internet access account. The only clue I have to my neighbors' usage is the delays that my traffic encounters on the way up and down the cable, and that only tells me how busy they are, not what sites they're visiting and which files they're downloading.

To obtain that level of information, I would have to use a Wi-Fi sniffer such as Air Pcap and hope their Wi-Fi networks are either completely unsecured or that they rely on an effectively useless cipher such as the deprecated *Wired Equivalent Privacy* standard known as WEP.

Anyone who uses a Wi-Fi network in a populated area without securing it with WPA or WPA2 is effectively sharing his personal web surfing and e-mail habits with any snoop who cares to hear them. This situation is intolerable to me, so I joined colleagues in the Wi-Fi Alliance in developing a system for quick and easy setup of secure Wi-Fi networks called Wireless Protected Setup or WPS. I hope all of you who use Wi-Fi understand that you're broadcasting your web surfing habits to anyone who cares to learn them if you haven't secured your networks. If you're forced to use an unsecured Wi-Fi network in exigent circumstances, you can provide yourself a measure of privacy by securing your e-mail connection with Transport Layer Security. Public Ethernet connections are also fundamentally insecure, as anyone connected to the same switch fabric you're connected to can easily capture your packets and examine them to his heart's content.

As a purely technical matter, there's no difference between the means that Wi-Fi engineers use to diagnose network problems and those used by snoopers on public Wi-Fi networks to steal passwords: the same packet capture tool can do both. But one activity is legitimate (and even necessary to the proper functioning of networks) and the other is not.

So my first recommendation to the committee is to **emphasize intent and behavior rather than technology** in its continuing efforts to protect communication privacy. Technologies are neither good nor bad, it's the uses we put them to that matter.

The Culture of Over-Sharing

Another threat to consumer privacy, and in my mind a much greater one, is what I'll call the Culture of Over-Sharing. With the advent of personal web sites, blogs, social networks, and Twitter, people are sharing information about themselves that would certainly make their grandparents blush. I follow a number of tech journalists on Twitter, and I can now tell you more details of their personal health, diets, and dating habits than about the stories they cover or the conferences they attend. I don't particularly care for this personal information, but it's a part of the package.

Stories abound about young people who've posted drunken party pictures of themselves while they were in college finding embarrassment, often costly, when they apply for jobs and have to explain their antics to Google-savvy recruiters. The Internet is a harsh mistress, and much of what happens there stays there, seemingly forever.

I've been operating a series of blogs on technology and politics since 1995, and recently have received a number of requests from past commenters to remove missives they posted to a blog a few years ago. One recent correspondent said his roommate had posted radical sentiment under his name (I have no way to verify one way or another,) and another admitted frankly to being young, reckless, and grammatically challenged when posting comments that he now feels make him less employable. So I've adopted a policy of removing older comments for any reason at all. The lesson that I draw from this is that **retention policies are critically important to privacy**. It's the nature of networks to disseminate information, public and otherwise, but the game doesn't change radically until past, present, and future are combined into large, searchable archives that holds us captive to our pasts forever. People, especially those who were young once, need to have the ability to reinvent themselves, and our culture of over-sharing combined with our massive Internet archives, is eroding it.

Consumer Education

I've alluded to consumer awareness, or the lack thereof already, but I'd like to emphasize it as there have been recent instances of inadvertent sharing. CNet News reports[12] that the Committee on Oversight has heard testimony on the following events:

- On February 28, 2009, a television station in Pittsburgh reported that the blueprints and avionics package for "Marine One," the President's helicopter, was made available on a P2P network by a defense contractor in Maryland.

- On February 26, 2009, the Today Show broadcast a segment on inadvertent P2P file sharing, reporting that social security numbers, more than 150,000 tax returns, 25,800 student loan applications, and nearly 626,000 credit reports were easily accessible on a P2P network.

- On February 23, 2009, a Dartmouth College professor published a paper reporting that over a two-week period he was able to search a P2P network and uncover tens of thousands of medical files containing names, addresses, and Social Security numbers for patients seeking treatment for conditions such as AIDS, cancer, and mental health problems

- On July 9, 2008, the Washington Post reported that an employee of an investment firm who allegedly used Lime Wire to trade music or movies inadvertently exposed the names, dates of birth, and social security numbers of about 2,000 of the firm's clients, including Supreme Court Justice Stephen Breyer. There have been reports alleging file sharing programs have been used for illegal purposes, such as to steal others' identities.

Technology always moves faster than regulation, and we want to keep it that way, but consumers need to be aware that some of the applications they run, particularly peer-to-peer file sharing applications, expose more information than they may want. It's unlikely that producers of Peer-to-Peer applications will be responsive to Congressional mandates

[12] Greg Sandoval, "Congress to Probe P2P Sites over Inadvertent Sharing," *CNet News*, April 21, 2009: http://news.cnet.com/8301-1023_3-10224080-93.html?part=rss&subj=news&tag=2547-1_3-0-20

of full disclosure; theirs is a quirky community with little regard for authority, but steps can be taken to make consumers aware of the dangers of inadvertent over-sharing.

Malware and Botnets

Perhaps the most significant threat to consumer privacy is deliberate identity theft. By now, this threat is well-understood: millions of computers worldwide are infected with viruses that put them under the effective control of the virus' creators. Infected computers, tied together in a huge *botnet*, are used to send Spam and to run key-loggers that steal personal information. The end produce is sent back to the controller where it's used for criminal purposes. It's suspected that some botnets may be controlled by foreign intelligence services as they're shown up in interesting places, such as the Dalai Lama of Tibet's offices in India. While Spam is in integral part of the Internet's e-mail system today, and will remain so as long as we don't adopt a system of user authentication as part of normal e-mail practice, efforts to mitigate its effects are impressive.

Spam fighters maintain a set of DNS Blacklists which squelch, by their estimation, some 81% of attempted Spam at the source, simply by checking the Internet Domain Name of the source networks against a list of known Spam networks. This is a very important function, but it raises the shackles of some privacy advocates, who see it as discriminatory and non-transparent. DNS Blacklists certainly do contain false positives from time to time, but they incorporate procedures for the removal of domains unfairly listed. The value of this kind of Spam mitigation is enormous, and it goes beyond the protection of consumer privacy: Spam has a considerable carbon footprint and contributes to global warming. According to a recent report by McAfee, Inc[13]:

- *An estimated worldwide total of 62 trillion spam emails were sent in 2008*
- *The average spam email causes emissions equivalent to 0.3 grams of carbon dioxide(CO2) per message*
- *Globally, annual spam energy use totals 33 billion kilowatt-hours (kWh), or 33 terawatt hours (TWh). That's equivalent to the electricity used in 2.4 million homes, with the same GHG emissions as 3.1 million passenger cars using two billion U.S. gallons of gasoline.*
- *Spam filtering saves 135 TWh of electricity per year. That's equivalent to 13 million cars off the road*
- *Much of the energy consumption associated with spam (nearly 80 percent) comes from end users deleting spam and searching for legitimate email (false positives). Spam filtering accounts for just 16 percent of spam-related energy use.*

Clearly, Spam mitigation is a social good. The Blacklist method isn't sufficient on its own; it's driven by intelligence about which e-mail messages are Spam and which aren't. This determination is made by a number of means, one of them human intelligence, but machines are part of the process as well. Mechanical recognition of Spam depends on a process of pattern matching e-mail against known contents of Spam currently in circulation in the Internet. Like anti-virus software, Spam detectors search for Spam

[13] *The Carbon Footprint of Email Spam Report*, McAfee Inc. and ICF International, http://img.en25.com/Web/McAfee/CarbonFootprint_28pg_web_REV.PDF, retrieved April 21, 2009.

signatures in ordinary e-mail, flagging or deleting suspect messages. This is in fact a very invasive process, one that can often cause legitimate messages to end up in user's spam folder or worse. But it's a system that Internet users embrace because its benefits far outweigh its drawbacks.

The lesson I suggest we should learn from Spam mitigation is to **examine mechanical processes for their practical benefits as well as their theoretical harm** to abstract notions of privacy, and to consider what our networking experience would be like without them. The damage to personal privacy inflicted by Spam signature searches has to be balanced against the greater harm that unchecked Spam inflicts. Similarly, an e-mail ethos based on personal identity rather than semi-anonymous access has benefits that are not lost on the architects of Internet e-mail. Future systems will surely be designed in more robust manner.

Traffic Engineering

Contrary to popular belief, the physical networks that carry Internet Protocol packets are not "stupid" networks. Most IP networks of significant size carry a combination of generic Internet traffic and private IP traffic that has to be delivered according to Service Level Agreements (SLAs) between end-user organizations and network carriers. Some SLAs are very stringent, allowing for as little as 2 milliseconds of latency (delay) between transmitter and receiver. In order to satisfy the needs of customers with varying SLAs, network operators buy network equipment that's capable of prioritizing packets. These systems depend on the ability to classify network flows[14] and to count packets per second over long periods of time. Their prioritization function interacts with accounting and policy functions to promote or demote specific flows depending on the customers standing in terms of volume and rate and his contract. Traffic engineering of this sort, generally using the MPLS[15] protocol which reduces the overhead of repetitive route lookup as the packet moves from one router to another, is at the heart of the modern Internet.

A simplified form of traffic engineering is now employed by Comcast on its residential broadband network to protect IP service from overload. When a link has been congested for a meaningful period of time, the system identifies heavy users of network resources (bandwidth.) Any of these users who've exceeded a meaningful threshold are placed in a lower priority category until the load they offer the network declines. This system is notably "protocol agnostic" as it treats all Internet applications the same: if a user is engaged in a large file transfer for a significant period of time (15 minutes or more) that places him in the low priority category, and if the user is also using Skype or some other VoIP service, his VoIP performance will suffer until he takes steps to curtail his downloading.

This system addresses one of the fundamental architectural shortcomings of the Internet, the absence of a per-user fairness system. This problem has been addressed in numerous

[14] A "flow" is a series of packets between a common source and destination.

[15] E. Rosen, A. Viswanathan, and R. Callon, *Multiprotocol Label Switching Architecture*, January 2001, IETF RFC 3031, http://tools.ietf.org/html/rfc3031

forms, and perhaps most clearly by Dr. Bob Briscoe, Chief Scientist at British Telecom Research[16]:

> *Resource allocation and accountability keep reappearing on every list of requirements for the Internet architecture. The reason we never resolve these issues is a broken idea of what the problem is. The applied research and standards communities are using completely unrealistic and impractical fairness criteria. The resulting mechanisms don't even allocate the right thing and they don't allocate it between the right entities. We explain as bluntly as we can that thinking about fairness mechanisms like TCP in terms of sharing out flow rates has no intellectual heritage from any concept of fairness in philosophy or social science, or indeed real life. Comparing flow rates should never again be used for claims of fairness in production networks. Instead, we should judge fairness mechanisms on how they share out the 'cost' of each user's actions on others.*

The Internet is a system built on the dynamic sharing of network bandwidth, but it lacks a general-purpose mechanism of allocating it across user accounts fairly. Because the Internet lacks this vital mechanism, it's necessary for network operators to supply it themselves, as they have since the first deployment of Internet Protocol in a wide-area network by Ford Aerospace in 1981.

While the network engineering community is acutely aware of the limitations of the Internet's architecture and protocol design, advocates for open access and related causes often gloss over this issue in their search for the perfect network. The network technician who discovered the original Comcast system for managing P2P by injecting TCP Reset packets complained that the user-volume-based system amounted to "discrimination based on user-history [sic][17]." If that's the case, it's a brief history, no more than 15 minutes long.

The lesson to be learned about traffic engineering is that the realities of business and the shortcomings of the Internet as a global system for multiple uses often collide with utopian desires for the more perfect network. In very real sense, the TCP/IP Internet remains a work in progress 35 years after it was proposed as a research network for the exclusive use of highly-trained network engineers, professors, and graduate students. It was somewhat unfortunate that it was pressed into service in a completely different role in the early 90s when the plethora of personal computers demanded interconnection. Compromises against ideals of network function are inevitable in this scenario, and should not automatically be judged failures simply because they violate abstract notions of network design that have never been more than pipe dreams.

[16] Bob Briscoe, *Flow-Rate Fairness: Dismantling a Religion*, http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/2020comms/refb/fair_ccr.pdf

[17] Robb Topolski, *Re: [p2pi] Follow-Up from Comcast Presentation*, e-mail to IETF P2PI, June 6, 2008. http://www.ietf.org/mail-archive/web/p2pi/current/msg00072.html

Standards bodies continue to address the Internet's need for continual improvement, and researchers are hard at work on several projects that would replace the current Internet with an improved network that reflects some of the knowledge we've gained in the last 35 years. In the meantime, I would urge the Committee **not to hold Internet operators to unrealistic standards**. Keeping the Internet running smoothly is a difficult task in the best of times, and any practice that has a plausible connection with this goal should be seen as constructive and responsible, even if it requires accounting for usage and acting accordingly. In the long run, traffic management systems that rely on accounting and prioritization are much friendlier to innovation than those that simply charge for usage.

Deep Packet Inspection

Discussions like this one inevitably come to Deep-Packet Inspection, that poorly-defined term that seems to portend something ominous ("it was a dark and stormy night for IP packets.") [18]

As I've endeavored to show, there are legitimate and illegitimate uses for most aspects of network technology, and this is no exception. If we recognize that much of the traffic on the Internet is digital piracy (it doesn't matter how much as long as we agree that it's significant,) we have to accept that some means of mitigation is appropriate, just as it is for Spam, viruses, and overload. The most effective means of piracy mitigation – other than jail time for the operators of piracy-enabling web sites and tracker services like The Pirate Bay – is a system in which piracy cops enter swarms of users downloading and sharing digital material in a manner contrary to law. This system doesn't rely on DPI, as it simply uses non-encrypted information made available by the pirates, perhaps inadvertently. Encryption doesn't make this any harder to do, as the fabric of P2P piracy is sharing known content with random partners across a network. These exchanges revolve around a content identifier know as a file "hash" which is computed across the entire range of a file. File hashes can be extracted from certain P2P transactions automatically, and these transactions can point the piracy cop toward trackers who may not have been known at the outset. Hence, DPI has a role, albeit a limited one, in piracy mitigation. As long as digital piracy is against the law, there has to be some accepted means of finding it and stopping it. This needn't involve door-to-door searches or trips to Guantanamo Bay, but it's not simply a matter of sitting on our hands and saying, as the founders of The Pirate Bay said after their conviction in a Swedish court, that anything easy to do should be legal[19].

[18] A great deal of the animosity against DPI seems to stem from the belief that a functional layering approach to communications regulation should replace the current model, which FCC Commissioner McDowell has described as "technology silos." The silo model is defective because it focuses on technologies rather than services, and breaks down in the face of the similarity between video, IP transport, and voice services delivered across multiple technologies. Unfortunately, the functional layering model simply rotates the silo 90 degrees, and retains multiple ambiguities due to the fact that networks often perform similar functions – such as retransmission and error detection – at multiple layers. The service and disclosure model currently used by the UK telecom regulator, Ofcom, is far superior to either.

[19] Owen Thomas, "Jail Time Shuts Down The Pirate Bay Joke Machine," *Valleywag*, Apr 17 2009. http://gawker.com/5216499/jail-time-shuts-down-the-pirate-bay-joke-machine

DPI can also be useful as a means of relaxing per-user quotas imposed by a fairness system, to better tune network service to application requirements. In a more perfect Internet – the one envisioned by the architect of the IP datagram's Type of Service field, the architects of Integrated Services, and the designers of Differentiated Services – applications should be able to communicate requirements to the network, and the network should do its best to meet them according to the service level that a user has purchased. It's for this reason that the Internet Protocol's header structure includes a field for signaling such requirements to the network. Unfortunately, in the transition from research to production network, this signaling was overlooked. Moving the Internet off the NSF Backbone and onto a mesh of private networks required the invention of a new protocol for service providers to communicate routing information with each other. This protocol – Border Gateway Protocol (BGP) – did not include a mechanism for attaching Quality of Service levels to routes. Private IP networks overcome this problem by adopting MPLS and using Ethernet VLANs, but the problem of communicating QoS levels in the public Internet remains unresolved. There is hope that a draft pending before the IETF's Inter-Domain Routing Working Group provides the solution[20]. The creator, a professor at the Chemnitz University of Technology, has tested his solution in number of public Internet exchanges in Europe and reserved the necessary numbers from ICANN.

In the meantime, the most effective way of determining application requirements is to examine streams and map them to QoS categories by their evident properties. Generally speaking, networks can provide the greatest utility if they can expend their most scarce resource, low-latency delivery, on the packets most in need of it. In the consumer scenario, these are VoIP packets. VoIP is a low-bandwidth application, generally requiring no more than 128 kilobits/second, and often much less. It's a stringent application in terms of delay, however, as it can't tolerate latencies greater than 150 milliseconds (thousands of seconds) from end to end. VoIP is generally recognized as a candidate for a network boost. P2P file sharing, on the other hand, is a candidate for demotion because it tends to use as much bandwidth as is available, and to do so for a very long time, often in the range of hours. Once an ISP has determined stream requirements, it can adjust its handling so as to provide rapid delivery for VoIP and economical delivery for P2P. This is simply a matter of assigning packets to appropriate SLAs within and without the ISP's network. If every interconnected part of the Internet doesn't immediately support such an extension of past functionality, there's no cause for alarm as some day most will. The intersection of technology, economics, and marketing is too compelling for any other outcome.

So there's no reason to fear the use of DPI for traffic engineering. There is no loss of personal privacy from such behavior, nor would its adoption drive the Internet into a posture that's less friendly to competition. If anything, **the ability of applications to select a transport service appropriate to their needs would be an enormous boon to developers** of either time-sensitive or volume-sensitive applications. They only suffer if all traffic has to be treated as if it were the same when it's clearly different.

[20] Thomas Martin Knoll, Simple Inter-AS CoS, March 9, 2009. http://www.ietf.org/proceedings/09mar/slides/idr-5.pdf

Tracking Cookies

One development that concerns me is the expanded use of tracking cookies to build dossiers of user behavior across the Internet. The most notorious current example is the Double-Click DART cookie[21] used by Google's AdSense program. The DART is a unique identifier placed in a user's computer by Google to track his or her movements around web sites that participate in the AdSense contextual advertising program. DART cookies as currently conceived are not especially evil – they simply allow advertisers to know how many times users have seen their ads on average, and which web sites are frequented by the same people - but there's something creepy about writing a blog post critical of Google and knowing that everyone who reads it essentially reports as much to the mother ship. Although the DART identifier is simply a random number with no particular connection to a discernable human being, the portion of the Internet's population who have both Gmail accounts and DART cookies certainly are potentially identifiable to anyone with sufficient access to Google's data base.

The prospect of ever-increasing dossiers of Internet users with information about who they are, where live, who their friends are, what blogs they read, and what trips they take is simply disturbing. While there is no evidence that this tracking data has yet been abused, it's simply a matter of time until a deranged Googler tracks an ex-girlfriend or an over-ambitious product manager applies some artificial intelligence to predict what we will buy that we didn't even know we wanted.

I have no particular recommendation regarding tracking cookies and the related dossiers but for the Committee to keep an eye on the way they're used and on the lookout for feature creep. All collectors of information seem to share the attitude that if a little bit of information is good, a lot is better, and all information tends to leak over time.

Conclusion

The most effective means of monitoring consumer behavior is a well-placed virus, and failing that it's a system of web tracking with a persistent cookie linked to a personal account. A number of technologies with primarily beneficial uses have been demonized for eroding privacy, often unfairly. The greatest threats to consumer privacy are not technologies – we're awash in technology – but business models that depend on the bartering of personal information. The Internet is unfortunately surrounded and permeated by an "information wants to be free" ethos in which advertising is the key source of revenue for the providers of application and content-level services. This business model inevitably collides with personal privacy concerns, and needs to be constantly monitored. I fear the only way to ensure robust protection for personal privacy in the long run is to replace the open-access, advertising-supported business model with one in which we pay for content and services. Given the strength of the Internet's now well-established tradition of pushing ads into and alongside practically everything that we see, this is not going to be an easy transition, if it's to happen at all. But as long as personal information is the coin of the realm, it will be harvested, archived, and bartered.

[21] http://www.doubleclick.com/privacy/faq.aspx

Thank you for your kind attention,

Richard Bennett
BroadbandPolitics.com

Originally published by the House of Representatives Energy and Commerce Committee.

Appendix F: A Google monopoly today means packet snooping tomorrow: A plan to protect our privacy

This is an opinion column I wrote for The Register, the leading European technology news site, on June 29, 2009. It covers the implications of the April hearing in the House, arguing that regulations on data collection are less important to consumers than regulations on data protection and resale. The headline prediction is prescient, of course. Available at http://www.theregister.co.uk/2009/06/29/bennett_google_privacy/

A Google monopoly today means packet snooping tomorrow: A plan to protect our privacy

Now that America's lawmakers have repaired the world economy, they can turn their attention to more mundane matters, such as saving the Internet.

There's an inherent conflict between traditional notions of personal privacy and the Internet's emerging goldmine, targeted advertising. Other than the subscription fees that carriers collect for access to the Internet itself, the only reliable revenue stream the 'Net has ever generated is ad sales, which mostly depend on the advertiser having knowledge of the consumer's tastes and interests.

Google's targeted advertising program AdSense is even more intrusive than the controversial Phorm and NebuAd systems. For example, Gmail scans your personal communication for keywords - there is no opt-out, and using a secure tunnel is no protection. More recently, Google has stepped up the aggressiveness of its program by shifting the tracking cookie used by AdSense from an opt-in to an opt-out system of consent, where opting-out requires arcane knowledge on the part of the consumer.

The tension between privacy and revenue took center stage in a House Subcommittee on Communications, Technology, and the Internet hearing on Internet privacy at which I was a witness recently. The new chairman, Rick Boucher, intends to conduct a series of hearings around a privacy bill he's promised to introduce later in the session, the next of which will include actual ad merchants, such as Google and Yahoo.

No major American ISP is currently using DPI to track consumer behaviour, and the web trackers would prefer it remains that way. The practical implication of the current state of play would have Google gaining a functional monopoly on targeted advertising in the very near future, at which point we might reasonably expect Congress to beg ISPs to start using DPI to track consumer behaviour.

Instant Karma

As Scott McNealy and others have observed, there's precious little privacy on the Internet. I was reminded of this by the author of one of the first Internet RFCs on my flight to DC. But that doesn't prevent Google's champions from using the privacy canard to preserve the status quo. Rep. Anna Eshoo (D, Google) tried to skewer me before the committee because of a remark in my written testimony on the conflict between privacy and targeted advertising - I suggested that the only way to ensure personal privacy in the long term is for users to pay for content and services. The threat to privacy isn't technical. It is a consequence of the Internet's business model.

Eshoo quoted one of my sentences, calling it a modern day "Modest Proposal," and asked the fire-breathing privacy advocates what they thought about it. The answer she got set her back on her heels, as the only witness to answer, EPIC chairman Marc Rotenberg, took the point even further, warning that the growth of unfettered advertising would come to have a corrupting effect on publishing itself, leading to a credibility meltdown of sorts. Score 1-nil to the geek.

The Kumbaya Moment

While the hearing started on shaky legs, it was apparent toward the end that there's considerable agreement that a legal framework for personal privacy needs to be created that covers all the technical bases.

Until now, the privacy debate has focused on particular ways of obtaining preference and stressed opt-in vs. opt-out. This approach is wrong-headed, as web spiders can extract more personal information from the Internet than DPI can. So the privacy problem actually needs to focus on what happens to dossiers of personal information that ad merchants own, regardless of how the information was obtained.

The new consensus dictates that the key issues are the protection of archived information from abuse, consumer notification about what's held by whom and how it's used, and the ability to have archived information erased. In the course of the discussion I suggested that consumers need periodic reminders of which services are building databases on their behaviour and the ability to have them erased. This notion found favour with the committee and the other witnesses.

While Washington continues to host fanatics on both sides of the policy spectrum, the current mood is one of pragmatism and regulatory restraint. While Obama Administration figure Susan Crawford and members of Congress with close ties to Google (primarily Silicon Valley congresswomen Lofgren and Eshoo) continue to promote wild-eyed, Utopian notions of net neutrality that simply protect the search monopoly's position, my sense is that they're outnumbered by pragmatists who would be pleased to allow a lightly-regulated market and the public relations machinery of the public interest organizations to correct egregious practices wherever they're found.

How long this fit of temporary sanity will persist in Washington is anyone's guess, but for the moment there's not much to worry about on the banks of the Potomac.®

Published in The Register, London.